Oxford University Computing Laboratory

Enforcement of Constraints on XML Streams

Full-time Research Assistant
Grade 7: Salary £27,466 - £33,780 p.a.

Further particulars

The project concerns the development of *runtime monitors* which check for data quality violations within XML feeds and react to quality violations: reactions include terminating processing of the stream, dropping portions of the payload, or attempting a repair. We are particularly interested in XML feeds that arise in the context of secure message-processing. The project lies at the intersection of several topics – runtime monitoring, secure messaging, XML processing, and integrity-constraint validation

Integrity constraints list sanity conditions on data. For XML and other tree-like data we will consider schema-based constraints – given via a DTD or XML schema for the XML payload being transmitted -- as well as user-defined (i.e. application-dependent) constraints, which may be either static or dynamic. One constraint language of interest for user-defined constraints is the w3c's XPath language. XPath can be applied directly to streams representing a single XML document or message directly; it can also be applied to streams representing a sequence of documents or other hierarchical structures, by considering these as a single incomplete forest.

One basic issue concerns the ability to create efficient stream processors for XPath constraints. Prior results have shown that this is possible for large fragments of the navigational core of XPath; the technique used is to translate integrity constraints into symbolically represented automata, for example, automata based on BDDs. Despite this, the ability to perform efficient constraint validation is not well-understood, either theoretically or in practice. On the theoretical side, a good characterization of exactly which constraints can be processed with a tractable space footprint is still lacking, both for navigational XPath and its extensions with recursion. On the practical side, to catch validation errors as quickly as possible requires finding ways to optimize symbolically-represented automata.

BDD-based techniques have shown promise, but apply only to simple schemas (e.g. DTDs) and to user-defined constraints based on navigational XPath. Moving beyond these will require very different techniques. For many schema-based constraints - for example, XML keys and foreign keys - and for non-navigational XPath, it is not even clear what a reasonable model of efficient constraint-processing is - clearly classical automaton models are insufficient.

Another critical question concerns the ability to react to constraint violations. How can a feed be modified in streaming fashion to ensure that constraints are never violated?   From the

theoretical point of view, it will be necessary to extend results on constraint repair to the context of XML feeds.   From the practical point of view, it will be necessary to define a policy language for reacting to constraint violations

The aim of the project is to extend the theory of constraint processing in the XML context, and to develop a prototypical implementation that can efficiently deal with rich sets of constraints on large numbers of simultaneous message feeds.

The postdoctoral assistant will work with two doctoral students, principal investigator Prof. Michael Benedikt, and visiting researcher Dr. Alan Jeffrey of Bell Laboratories in:

- Investigating theoretical issues in XML constraint processing, including:

    o Devising techniques to prove bounds on the size of stream-processors for XPath constraints.

    o Investigating optimisation techniques for BDD-based XML stream processors

    o Extending the theory of efficient streaming validation to deal with both aggregate constraints and keys.

- Developing both policy specification languages and implementation techniques for repairing constraint violations.

- Developing a full constraint management system, based on the above mentioned theoretical results, and to conduct an empirical performance evaluation.

Selection criteria

Essential:

A first degree in Computer Science or related discipline.

Applicants should have, or be about to obtain, a doctorate in computer science or a related discipline

Have a research record in one or more of the areas of: database management; database theory and/or automata theory; automated verification/testing; programming languages.

Programming experience, preferably in Java or C++.

Interests or capabilities in XML processing.

Ability to give technical presentations and prepare reports/papers for publication.

A willingness to collaborate with others and work effectively as a member of a team.

Desirable:

Experience in the implementation of algorithms for BDD manipulation.

Experience of working in collaborative or interdisciplinary environments.

Salary and Benefits

Salary will be on the University grade 7 scale (currently £27,466 - £33,780 p.a). The post is available from January 2009, will be of three years duration, is pensionable and includes an annual leave entitlement of 38 days per year, inclusive of public holidays and university closed periods.

Application Procedure

Applications should be in the form of a letter of application (clearly stating the post title) setting out how the candidate meets the selection criteria, and supported by a full curriculum vitae, together with the names and addresses of two referees. These should preferably be emailed (most formats accepted) to **job20@comlab.ox.ac.uk** or alternatively, posted to: The Administrator, Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD to arrive by the closing date of **Friday 1 November 2008.   Applications received after this time will not be considered.**

Candidates must also ask their referees to consider these further particulars and email their reference directly to job20@comlab.ox.ac.uk or, alternatively, post or fax it to the above address (fax (+44 1865 283532) such that the reference arrives by, or shortly after, the closing date.

It is expected that interviews will be held during the week commencing 1 December 2008.

The policy and practice of the University of Oxford require that all staff are offered equal opportunities within employment. Entry into employment with the University and progression within employment will be determined only by personal merit and the application of criteria which are related to the duties of each particular post and the relevant salary structure. In all cases, ability to perform the job will be the primary consideration. Subject to statutory provisions, no applicant or member of staff will be treated less favourably than another because of his or her age, sex, marital or civil partnership status, sexual orientation, religion or belief, racial group or disability.